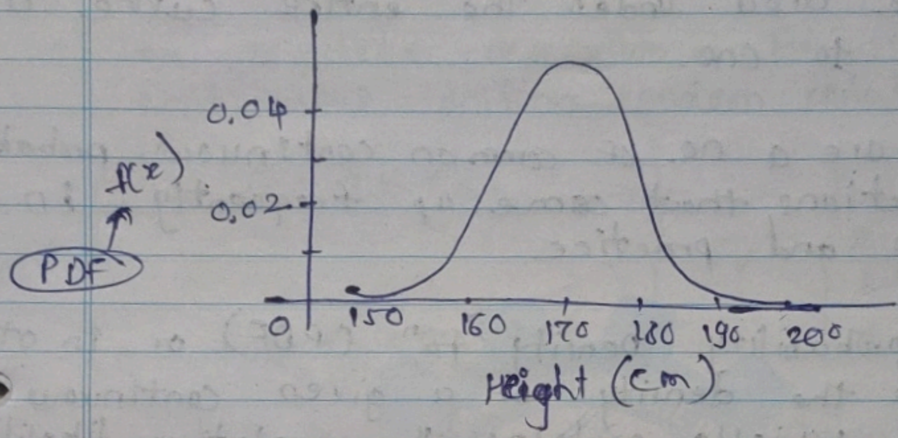
B.M.I.T.

* Contineous Probability Distribution *

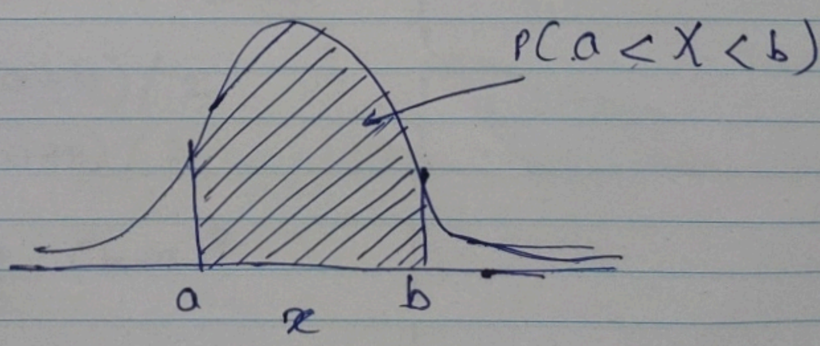
e.g ① Distribution of height of adult canadian males



- We cannot model continuous random variables with the same methods we used for discrete random variables.

- We model a continuous random variable with a curve f(x), called a probability density function (pdf)

- f(x) represents the height of the curve at point x.
- For continuous random variables probabilities are areas under the curve.



$P(X=a) = 0$, e.g: $P(X = 3.12) = 0$

- For any contineous probability distribution:
  - $f(x) \geq 0$ for all x
  - The area under the entire curve is equal to one.

- There are a no. of common continuous probability distributions that come up frequently in theory and practice.

- The probability Density fun (PDF) or in other words the density of a given continuous random variable explains the relative likelihood or probability for our continuous random variable to take on a given value in an interval.

$$f(x) = \frac{dF(x)}{dx} \quad \text{or} \quad F'(x)$$

- Probability density fun has two basic properties
  $f(x) >= 0$ for all real x

$$\int_{-\infty}^{\infty} f(u)\,du = 1$$

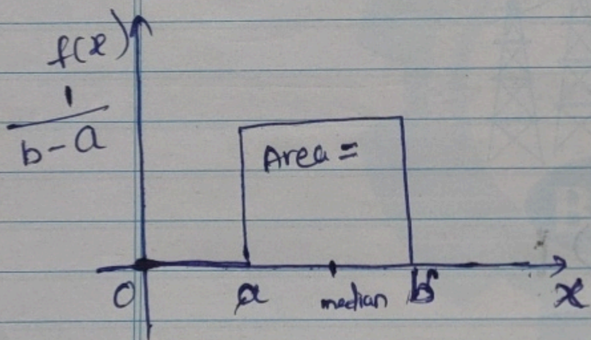
* Types of Continuous Distribution *

(1) Continuous Uniform Distribution :-

- The complete description of PDF and CDF of a continuous uniform random variable is as follows:

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{for } a \le x \le b \\ 0 & \text{for } x<a \text{ and } x<b \to \text{otherwise} \end{cases}$$

probability density fun



$$f(x) = \frac{1}{b-a}$$

where, $a <= x \le b$ is a continuous uniform random variable

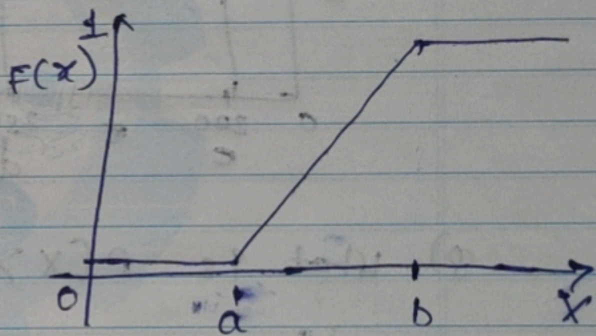
Median $= (a+b)/2$

$\mu = (a+b)/2$

$\sigma^2 = \frac{1}{12}(b-a)^2$

$$F(x) = \begin{cases} 0 & \text{for } x<a \\ \frac{x-a}{b-a} & \text{for } a \le x \le b \\ 1 & \text{for } x>b \end{cases}$$

Cumulative distn fun



$$F(x) = \frac{(x-a)}{(b-a)}$$

where, $a <= x \le b$

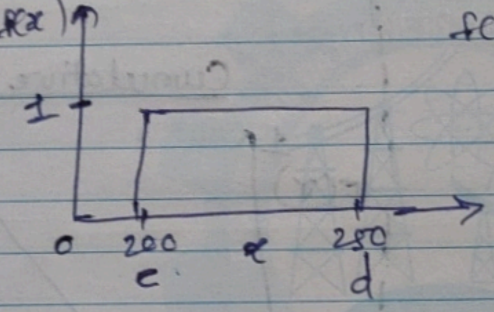
- For continuous probability distribution, finding area's usually requires integration
- But for the uniform distribution, areas under the curve are simply rectangles

q). E.g) Suppose X is a random variable that has a uniform distribution with $c = 200$ and $d = 250$

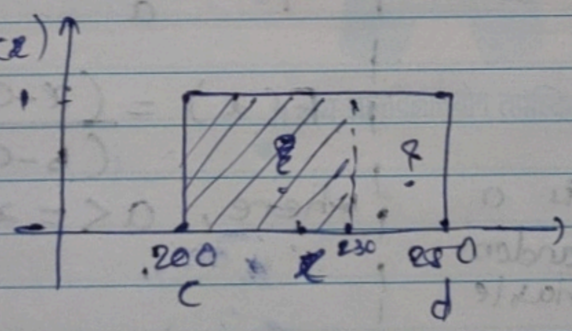
q. what is $f(x)$?

$$f(x) = \frac{1}{d-c} = \frac{1}{250-200} = \frac{1}{50} \quad \text{for } 200 \leq x \leq 250$$

$f(x) = 0$ ... for otherwise



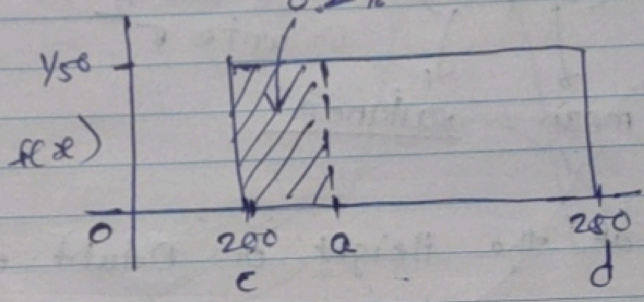
q). What is $P(X > 230)$?



$$f(x) = \frac{1}{d-c}$$

$$P(X > 230) = (250 - 230) \cdot \frac{1}{50} = 0.4$$

E.g What is the 20^{th} percentile of this distribution



$$(a - 200)\frac{1}{50} = 0.2$$

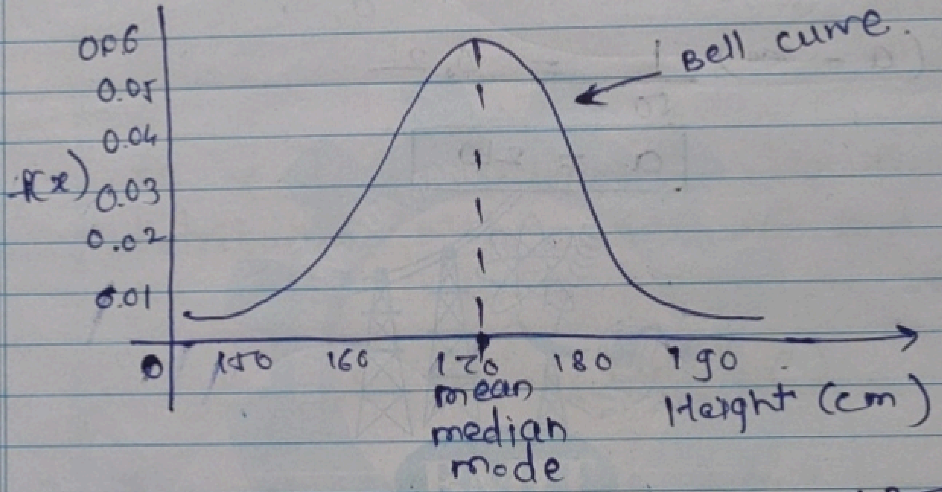
$$\boxed{a = 210}$$

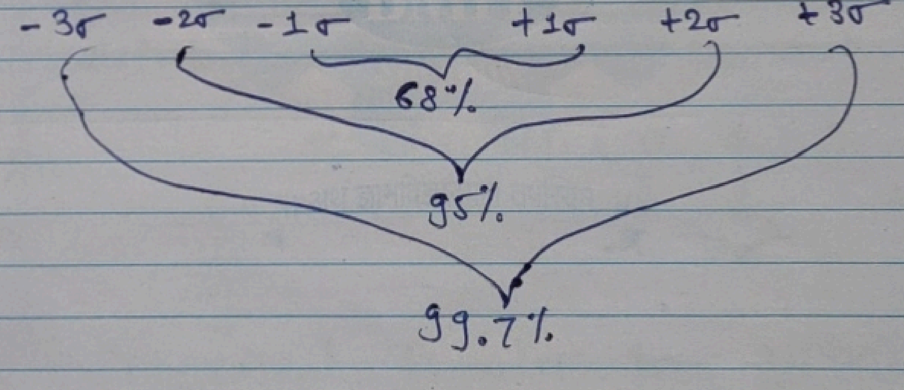
② Normal / Gaussian Distribution

- $X \sim G.D(\mu, \sigma)$ → we can write σ^2

X → Random variable; μ → mean; σ → variance

e.g. Distribution of the Height of Adult canadian males:





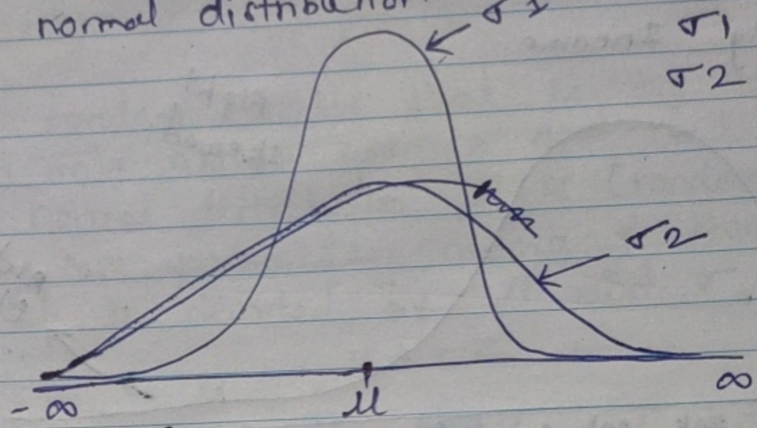
- The probability density function (PDF):

$$f(x) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{1}{2\sigma^2}(x-\mu)^2}$$

for $-\infty < x < \infty$

$$z = \frac{x - \mu}{\sigma}$$

Two normal distribution



$\sigma 1$
$\sigma 2 = 2\sigma 1$

- If x is a random variable that has a normal distribution with mean μ and variance σ^2, we write this as

$$X \sim N.D.(\mu, \sigma^2)$$

- The standard normal distribution is a normal distribution with mean = 0 and variance = 1

$$Z \sim N.D(\underset{\mu}{\underbrace{0}}, \underset{\sigma^2}{\underbrace{1}})$$

* Emperical Formulae in N/G.D

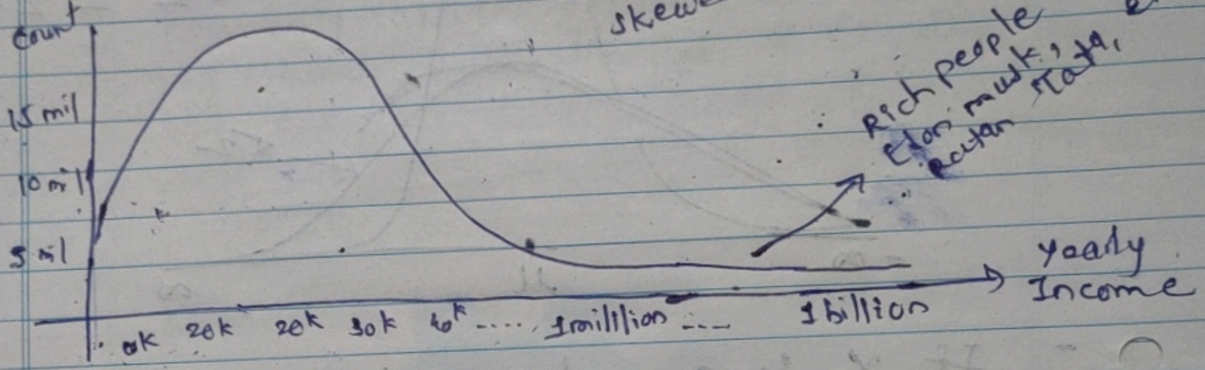
$$P(\mu - \sigma \leq x \leq \mu + \sigma) \approx 68\%$$

$$P(\mu - 2\sigma \leq x \leq \mu + 2\sigma) \approx 95\%$$

$$P(\mu - 3\sigma \leq x \leq \mu + 3\sigma) \approx 99.7\%$$

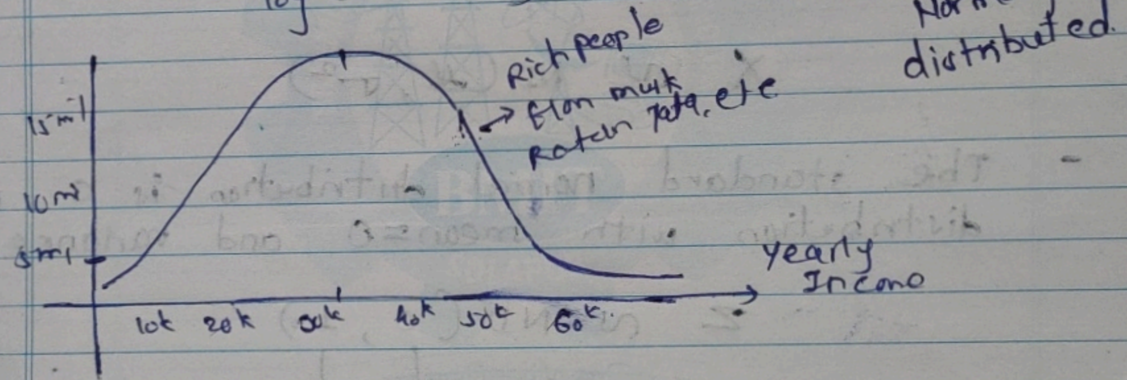
③ Log Normal Distribution

e.g. Yearly Income



⇓

log (Income)



- If we get a normal distribution by applying a log function to a dataset then dataset is log normally distributed.

- Random variable X is log normally distributed, $\ln(X) = W$ i.e. the natural logrithm of X normally distributed.

$$\ln(X) = W$$

$$\log_e(X) = W, \text{ is also equal to } e^W = X$$

Properties of Log Normal distribution

- The random variable that is log normally distributed, can only assign positive and real values in it.
- Log normal distribution of x (random variable) has two parameters mean & standard deviation which is denoted by μ and σ respectively.

$$x = e^{\mu + \sigma z}$$

where, z is a standard normal distribution.